RANDOM MOMENT SAMPLING TO ESTIMATE ALLOCATION OF WORK EFFORT

Richard Bolstein, George Mason University

1. INTRODUCTION

The proper determination of federal, state, and local share of administrative costs of operating a State Social Services Department is a complex and time-consuming task. Since cost sharing formulas and funding sources differ by type of service provided, the first step in this process is to accurately allocate staff time to the various types of programs. The state of Virginia currently uses a case-count method along with 'caseloads standards' which assign an average number of 'earned hours' to each type of program and activity. The total work effort for a period of time (month or quarter) allocated to a particular program is obtained by multiplying the number of cases for that program, broken down by 'activity' (initial determination, re-determination, fraud, etc.), by the appropriate standard. The problem with this method is in keeping the 'standards' updated, a difficult, time-consuming task (even if a sampling approach is used). Most other states use some type of labor-intensive approach which requires complete case reviews and/or logbooks kept by the staff. Recently, the U.S. Department of Health and Human Services (HHS) has urged states to adopt a work-sampling approach to this problem. This method, known as random moment sampling (RMS), intercepts workers at random 'time moments' and records what program and activity they are engaged in at that moment. Statistical methods can be used not only to estimate the allocation of total work time to each program, but also to provide error bounds for these estimates. The latter cannot be done with the 'complete-review' methods since the bias inherent in those methods is not computable. Furthermore, RMS is much less labor-intensive than the 'complete-review' methods. As a result, several states have adopted the RMS method over the past seven years. Most of these use a simple random sample of moments as suggested by HHS, although the structure of the social service system in some states requires a more complex sampling plan. Michigan uses a two-stage cluster design where the clusters are local welfare agencies which are chosen with probability proportional to size. Each worker in the selected clusters is interviewed at four random moments in the quarter. South Carolina stratifies by worker classification, whereas Pennsylvania and Tennessee employ systematic sampling. In devising a plan for the Virginia Department of Social Services (VDSS), the author was led to stratify the sample according to groups of agencies which are, to some degree, homogeneous with regard to the types of programs administered. The description of this RMS plan for VDSS is the subject of this paper.

The literature on random moment sampling is virtually non-existent in statistical journals. As the technique was originally applied to the textile and other manufacturing industries, it made its way into the industrial engineering and industrial psychology literature, usually under names like 'work-sampling' and 'time studies' (see Niebal [4]). One goal of writing this paper is to call attention to some interesting problems that arise with this technique in estimation and non-response adjustment that may be of interest to statisticians.

2. PROBLEM FORMULATION

Welfare programs in Virginia are carried out by 124 local agencies. These agencies have two primary functions:

(1) Determine client eligibility for financial assistance programs, and
(2) Provide or arrange for social services.

Most employees are involved with just one of these functions, and are classified as <u>eligibility</u> (or <u>benefit</u>) <u>program workers</u> and <u>service program workers</u> respectively. There are also some 'generic' workers who are involved with both functions.

The central problem is to determine the federal, state, and local share of administrative costs (worker salaries) to carry out these functions. The shares differ by the type of program within each function, as shown in Table 1 below. Although shares are the same for some of the programs shown in the table, separate funding sources require breakout of costs for each program listed.

If workers were dedicated to individual programs, it would be a simple matter to allocate costs since, for example, the total salaries of all ADC eligibility workers would be charged according to Table 1. As this is not the case, the next best approach is to obtain the proportion P_i of total worktime spent on program i. If F_i denotes the federal share of program i in Table 1 and C the total administrative costs for the time period in question, then it is natural to assign as the federal share of total costs the quantity

$$(\sum F_i P_i)C$$

The problem is to estimate the unkowns P_i. Costs are computed separately for eligibility and service functions (see [5]).
Workers are considered dedicated to that function which absorbs at least 80% of their time. For sampling purposes, generic (i.e. non-dedicated) workers are assigned to that function which occupies the majority of their time. However, a sampled moment is assigned to the function in which it belongs regardless of the function assigned to the worker. This happens infrequently and has an insignificant effect on sample size. Thus, for the function under consideration, P_i will denote the proportion of total work-time for that function which is spent on program i, and C will denote the total administrative costs of that function.

3. SAMPLE FRAME

To estimate the P_i, workers are intercepted at random times to find out what activity they are engaged in at those moments. This is the technique of random moment sampling (RMS). Thus, the <u>units of analysis</u> are <u>worker-moments</u>. Procedural difficulties arise if 'moment' is treated as a continuous variable since it is essential that the worker be intercepted at the precise moment selected. It is customary in RMS to divide the workday into equal time intervals and identify each interval with a specific 'moment' such as its' first minute. The common length of the time intervals should be at least as short as the shortest period a worker could spend on any one program (Chapanis [3]), which was judged to be five minutes in VDSS. The typical workday in VDSS is 8.5 to 9 hours long including an hour lunchbreak. However, hours are flexible with regard to start and finish time and length of lunchbreak and local agencies set their own requirements for number of hours worked. In order to insure that all workers have the same probability of inclusion in the sample, a standard workday of 7:30 A.M. to 6:00 P.M. was initially chosen, but later reduced after a pilot test to 7:30-5:30 for eligibility and 8-5 for services. Hence there are a total of 120 five-minute intervals per day for eligibility (108 for services) with the first minute of each interval designated as the corresponding moment. The number of eligibility moments M in a quarter is the number of workdays in the quarter (typically 62) multiplied by 120. If N denotes the number of eligibility workers in the quarter, then the eligibility sample frame consists of MN worker-moments. The services sample frame is constructed analogously.

Of course, not all worker-moments will correspond to program activity since the worker may be on a lunch or other break, annual or sick leave, not at work due to flex-time, or involved in some non-program specific activity such as training or other administrative duty. It is useful to define such non-program categories as 'leave', 'administration', and 'flex-time', since they can be used to help monitor the sampling process. For example, flex-time should account for 10 to 15% of worker-moments and hence if the sample result differs significantly from this there may be an error in the coding or interviewing process.Every worker-moment is either a program moment or one of the three non-program moments just defined. Denote by Q_i the proportion of the MN population worker-moments that fall in category i, and let Q denote the proportion of program moments. Then,

$$Q = \sum{}' Q_i$$

where the prime means that the sum is over all indices i corresponding to programs (from Table 1, there are 12 programs under the eligibility function and 9 under the services function). We also have

$$P_i = Q_i/Q$$

for all program categories i. Lower case letters will be used for the corresponding quantities computed from a sample. Recall that in a simple random sample of n worker - moments, q_i and $p_i = q_i/q$ are unbiased estimates of Q_i and P_i with estimated variances $q_i(1-q_i)/n$ and $p_i(1-p_i)/(qn)$ respectively ([2, Ch.2]).

4. SAMPLE DESIGN

A simple random sample is not appropriate for this problem for several reasons, including:
(1) Local agencies vary in size from as little as three employees to three-hundred. Thus, many small agencies would not appear in the sample, an important consideration in determining local reimbursement. Additionally, there are indications that small agencies have a different program mix than large agencies.
(2) Rare programs, those which account for less than 1 % of the work-effort, may not appear in even a moderately large sample. For example, according to the 'caseloads standards' method currently used in the State (see Section 1), Subsidized Adoption accounts for only one-tenth of one percent of the service program workload. Indeed, in a pilot sample of size 2400 no such moment was encountered.
(3) Some programs occur primarily in only a few agencies. For instance, the Refugee programs are almost non-existent outside Northern Virginia.

Most of these problems can be handled by appropriate stratification. The agencies in Northern Virginia involved with the refugee program make up one stratum, the large agencies of the Tidewater area in the southeast constitute a second, Richmond together with some other mid-sized agencies in the central part of the state which handle much of the 'rare' subsidized adoption cases form a third stratum, and the remaining agencies scattered throughout the state, which are primarily small and rural, make up the fourth and final stratum.

Not all of the 'rare' programs can be handled by geographic stratification. If some worker classifications could be identified which are more likely to be involved with such a program, that could provide an effective stratification. No such classifications have yet been identified in VDSS. However, one method for dealing with this problem is to combine samples from previous quarters for estimation of these programs, and then normalize the estimates of the other programs as computed from the present sample.

To summarize, the RMS design for VDSS is a geographically stratified sample with four strata and simple random sampling of worker-moments within strata.

5. DATA COLLECTION

An observation form was designed to be filled out by the worker and his or her supervisor at the prescribed moment (and required both signatures). The supervisor received a schedule of worker-moments about one week in advance. Workers did not know in advance when or if they were to be interviewed.

To obtain reliable data with a minimum of effort, the observation form should be short and clear but contain enough information to be able to assess the accuracy of the responses. In addition, it must provide information to help determine the causes of inaccuracies and missed observations (non-responses) so they can be remedied by training. For example, it frequently happened that the interview did not take place at the prescribed moment but was recalled at a later time (see Section 6). The observation form not only records the function and activity (program or non-program) of the worker, but also documents the length of recall (same day, or following day or later) and who was responsible for the missed observation (worker unavailable due to field work, or supervisor unavailable due to a meeting or just forgetfulness). The non-program categories were included on the form in some detail since, as mentioned in Section 3, they also provide a check. Thus, 'leave' is decomposed into four parts: annual, sick, lunch/break, other. The observation form itself is not shown here due to space limitations, but see [1].

6. PILOT TEST

A pilot test based on this design was conducted in the October-December quarter of 1985. It included 14 of the State's 124 agencies representing all four strata. Equal sample sizes were chosen for each stratum since results were to be compared by stratum. These were set at 800 for the eligibility sample and 600 for the services sample so that HHS precision requirements would be met ([5]). The sample size calculations were made from standard formulas for stratified samples (Cochran [2]) using estimates of proportions from a preliminary pilot study in June-August of 1985.

A total of 3009 eligibility and 2230 services observations were obtained from the pulled samples of 3200 and 2400 reflecting 94% and 93% response rates respectively. Table 2 below provides a convenient way of assessing the overall percentage of program strikes and timeliness of observations for the eligibility sample. Note that only half the samples resulted in program strikes. Part of this can be explained by the fact that the percent of program strikes prior to 8 A.M. and after 5 P.M. dropped to about 10%. (Subsequent analysis showed that the allocation of these extremal strikes to the various programs did not deviate significantly from the allocation of the entire sample. Therefore, the 8-5 workday is being used for the current statewide study and program strikes are running about 55%.) Note too that the lunch/break percentages seem low since one would expect about 1.5 hours out of 10 or 15%. It is quite possible that some workers are reluctant to report 'being on break'.

Late reports pose a problem for the validity of the data. However, it is clear that not all observations occur on time. The timeliness of observations was subdivided into three main categories: those that occurred on time, those that were recalled later the same day, and those that were recalled the following day or later. Fortunately, over 85% of all program observations were completed within the same day. Note that although non-program observations had a higher percentage of on-time completion, the proportion completed within the same day was no better than the program observations.

In order to try and reduce late observations, the reason for lateness was reported. Contrary to expectation, in the eligibility sample the observer was unavailable more frequently than the worker. In the services sample unavailability was about equally divided. Followup meetings with local supervisors indicated that the major problem was simply not remembering to make the observations on time. The use of alarm clocks and/or the designation of a single observer at a local agency for a day or week was recommended as a possible solution.

7. ADJUSTMENT FOR MISSING WORKERS

During the pilot study, a group of workers in one stratum were inadvertently omitted from the sample frame. Omissions can also occur naturally if new workers are hired during the quarter after the sample was drawn. Thus, the within strata estimates must be adjusted in such cases before making overall estimates. In this section we focus on a single stratum in which this problem occurred, present a method of adjustment, and apply it to the pilot test. Let L denote the number of workers omitted from the sample frame and N the number included, so that N + L is the total number of workers. (Both N and L may be fractional since not all workers are employed the full quarter.) If q_i is the sample estimate of the proportion of moments in the sampled population allocated to category i, then the adjusted estimate of Q_i is

$$q_{ai} = [N/(N + L)]q_i + [L/(N + L)]R_i$$

where R_i is the (unknown) proportion of the ML unsampled population moments allocated to category i. Thus, the stratum under consideration is decomposed into two substrata consisting of the sampled and unsampled populations. In general, the unsampled population can be broken down into several substrata to facilitate the estimation (actually, educated guesstimation) of the R_i.

In the pilot study, all L missing eligibility workers came from one agency and were dedicated to one program (fuel). At that agency, the normal workday was 8.5 hours including one hour for lunch and an estimated half-hour for breaks. As they were temporary workers , no vacation time was available and sick leave was considered minimal. Since the eligibility sample workday was ten hours long, we assigned the values

$$R_{flex} = .15, \quad R_{leave} = .15, \quad R_{fuel} = .70$$

and $R_i = 0$ otherwise. Such guesstimates will be refined by isolating the observations of fuel workers that are sampled.

The adjusted estimate of the proportion P_i of program moments that fall in category i is

$$p_{ai} = q_{ai}/(\sum_j{}' q_{aj}) = [Nq_i+LR_i]/[Nq+LR]$$

where $R = \sum_j{}' R_j$.

8. ESTIMATION

The formula for an unbiased estimate of the statewide proportion Q_i in a stratified sample is (Cochran, [2,Ch. 5])

$$q_i = \sum_h (W_h)q_{hi}$$

where q_{hi} is the estimate in stratum h, and W_h is the proportion of the population in stratum h. If M is the number of moments in the quarter, N_h the number of workers in stratum h, and N the total number of workers in the population, then

$$W_h = MN_h/MN = N_h/N$$

An unbiased estimate of the variance of q_i is

$$V(q_i) = \sum_h (W_h)^2 (q_{hi})(1-q_{hi})/n_h$$

where n_h is the size of the sample from stratum h (that is, the number of worker-moments sampled from stratum h).

Estimation of the P_i is more complex. We present two methods.

Method I. Analogy with simple random sampling suggests setting

$$p_i = q_i/q$$

where

$$q = \sum_j{}' q_j$$

is an estimate of the proportion Q of work moments that correspond to program activity (we call these program moments). If we let

$$q^{(h)} = \sum_j{}' q_{hj}$$

denote the sample proportion of program moments in stratum h, then

$$q = \sum_j{}' \sum_h W_h q_{hj} = \sum_h W_h (\sum_j{}' q_{hj}) = \sum_h W_h q^{(h)}$$

so that

$$p_i = \sum_h (N_h q_{hi}) / (\sum_h N_h q^{(h)})$$

is seen to be a combined ratio estimate (Cochran [2, p. 165]). The variance of p_i can be written approximately as

$V(p_i) =$

$$\sum_h [W_h{}^2/(Q^2 n_h)][Q_{hi}(1-Q_{hi}) + P_i{}^2 Q^{(h)}(1-Q^{(h)}) - 2P_i Q_{hi}(1-Q^{(h)})]$$

where $Q^{(h)}$ and Q_{hi} are respectively the stratum h population proportion of work moments that are program moments and the proportion of work moments that are program i moments. These quantities and P_i are estimated from the sample to obtain an estimate of the variance. In large samples the bias of the ratio estimate is negligible if the coefficient of variation of $q^{(h)}$ is small (Cochran [2]).

Method II. The separate ratio estimate ([2,p.164]) of P_i can be written in the form

$$\hat{p}_i = (1/MNQ) \sum_h (q_{hi}/q^{(h)}) MN_h Q^{(h)}$$

$$= \sum_h (Q^{(h)}/Q) W_h p_{hi}$$

which now has the appearance of a stratified proportion estimate with unknown weights $(Q^{(h)}/Q)W_h$. If we assume that the proportion of moments that are program moments is the same in each stratum, then all $Q^{(h)} = Q$ and the weights become the familiar W_h. Thus, it seems reasonable to use the W_h as estimates of the true weights, in which case the estimate of P_i becomes

$$\hat{p}_i = \sum_h W_h p_{hi}$$

The variance and bias are given by

$$V(\hat{p}_i) = \sum_h W_h{}^2 p_{hi}(1-p_{hi})/[n_h q^{(h)}]$$

$$B_i = \sum_h W_h [1-Q^{(h)}/Q] P_{hi}$$

These quantities can be estimated from the sample. As the survey will be repeated on a continuing basis, improved estimates of the $Q^{(h)}$, and hence the weights can be obtained as a moving average from previous quarters to virtually eliminate the bias. However, even with the current estimate the bias turns out to be negligible in comparison with the variance in the pilot study. So how do the two different estimators compare? We partially illustrate with some data from the eligibility sample in the pilot study.

	STRATUM			
	1	2	3	4
n_h	783	751	796	663
W_h	.2596	.1622	.2694	.3088
$q^{(h)}$	.4700	.4727	.5340	.5284
$1-q^{(h)}/q$	.0706	.0653	-.0560	-.0449
q_{hi}(ADC)	.1264	.1691	.1809	.2055
p_{hi}(ADC)	.2690	.3577	.3388	.3890

The last row is just the fourth row divided by the second row. The estimate of the proportion of all moments that are ADC program strikes as computed from the first formula in this section is q_i = .1725 with variance .0000496. Also,

$$q = \sum W_h q^{(h)} = .5057$$

and

$$v(q) = \sum W_h{}^2 q^{(h)}(1-q^{(h)})/n_h = .0000887$$

So, $p_i = q_i/q = .3411$

whereas, $\hat{p}_i = \sum_h W_h p_{hi} = .3393$

So the two methods produce nearly identical estimates for the proportion of program moments allocated to ADC. The estimated variance of the first estimate turns out to be .000160. Since the coefficient of variation of q, which serves as an upper bound for the ratio of the square of the bias to the variance, is .0186, the bias of the composite ratio estimate is negligible ([2, p. 162]. The estimated variance of the second estimate is .000149 and the bias as computed from the table is b = -.000195. Hence, the square of the bias is .016 times the variance so is negligible in this case too. The relative design effect of Method I to Method II for ADC is 1.07,

TABLE 2

TIMELINESS OF THE ELIGIBILITY SAMPLE FOR THE DECEMBER 1985 QUARTER: UNWEIGHTED COUNTS AND PERCENTS

COUNT ROW PCT	ON TIME	SAME DAY WORKER	SAME DAY OBSERVER	NEXT DAY WORKER	NEXT DAY OBSERVER	ROW TOTAL
	899	129	302	27	149	1506
BENEFITS	59.7	8.6	20.1	1.8	9.9	50.0
	136	3	4	4	10	157
ANNUAL	86.6	1.9	2.5	2.5	6.4	5.2
	82	2	1	5	3	93
SICK LEAVE	88.2	2.2	1.1	5.4	3.2	3.1
	79	10	3	6	4	102
OTHER LEAVE	77.5	9.8	2.9	5.9	3.9	3.4
	213	27	21	6	14	281
LUNCH-BREAK	75.8	9.6	7.5	2.1	5.0	9.3
	145	17	30	9	22	223
ADMIN	65.0	7.6	13.5	4.0	9.9	7.4
	345	18	15	19	8	405
FLEXTIME	85.2	4.4	3.7	4.7	2.0	13.5
	230	1	3	0	8	242
VACANT	95.0	.4	1.2	0	3.3	8.0
COLUMN	2129	207	379	76	218	3009
TOTAL	70.8	6.9	12.6	2.5	7.2	100.0

indicating confidence intervals will be 3.6% wider under Method I. Similar results were obtained for the other program categories. A theoretical comparison of the methods will appear in another paper, but it seems experimentally that Method II is to be preferred because of its simplicity.

9. SUMMARY

Random Moment Sampling is a complex survey procedure which requires advanced methodology. Problems addressed include the handling of 'rare' subpopulations, assessing and improving the quality of data, adjusting for population elements omitted from the sample frame, and choosing among two estimation methods. Major issues not dealt with in this paper, due to time and space limitations, include elimination of zero cells and estimation of proportions for 'rare' programs, and variance estimation of the federal, state, and local share of costs.

REFERENCES

[1] R. Bolstein, G. Ballentine, J. Vaughan, "Implementing Random Moment Sampling in the Virginia Department of Social Services: Design and Analysis", Proceedings of the Twenty-Sixth annual Workshop of the National Association for Welfare Research and Statistics, 1986.

[2] W. Cochran, "Sampling Techniques", John Wiley & Sons, third edition, 1977.

[3] A. Chapanis, "Studying Work Activities in Office Systems", IBM Technical Report, 1976.

[4] B.W. Niebal, "Motion and Time Study", Richard D. Irwin, Inc., 1972.

[5] "A Guide for State and Local Government Public Assistance Agencies/Departments, Procedures for the Preparation and Submission of Cost Allocation Plans", U.S. Department of Health and Human Services, 1981.